



TITLE:

サポートベクターアルゴリズムに
対する切除平面法を用いた新解法
(数理最適化から見た「凸性の深み
,非凸性の魅惑」)

AUTHOR(S):

矢島, 安敏; 川副, 智司

CITATION:

矢島, 安敏 ...[et al]. サポートベクターアルゴリズムに対する切除平面法を用いた新解法
(数理最適化から見た「凸性の深み,非凸性の魅惑」). 数理解析研究所講究録 2004, 1349:
63-82

ISSUE DATE:

2004-01

URL:

<http://hdl.handle.net/2433/24861>

RIGHT:

サポートベクターアルゴリズムに対する切除平面法を用いた新解法

東京工業大学 経営工学専攻

矢島 安敏 (Yasutoshi Yajima)

川副 智司 (Satoshi Kawasoe)

Tokyo Institute of Technology

1 はじめに

Vapnik[10] によるサポートベクターマシン (SVM) を用いた判別法は、近年さまざまな分野の判別問題に応用され、非常に有力な手法と考えられるようになってきた。特に、文書の分類や画像の認識といった分野では、最も強力な手法の一つと考えられている。

SVM による判別は、従来の判別法と比べると最適化問題を解く必要があるなど計算が複雑なため、以前では大規模データに対しては適用できないと考えられていた。しかし、計算機能力の進歩と高速な特殊解法の登場で、近年では数十万データの問題に対しても実用的な計算時間で判別関数が求められるようになってきている。

本稿では、まず次節において、Vapnik による標準的な SVM の定式化について述べ、3 節では、2 クラスの分類問題に対する切除平面法を利用した解法について述べる。また 4 節では、提案する切除平面法によるアプローチを回帰問題へと適用する。

2 SVM による判別

2.1 SVM の定式化

N 属性のデータが M 個与えられている。各データ $j = 1, 2, \dots, M$ を N 次元空間 \mathbb{R}^N の点と考え、 N 次の行ベクトル A_j で表すこととする。また、各点 j には 2 値のラベル $y_j \in \{-1, +1\}$ が与えられている。このとき、ラベルの値にしたがって点を判別する 2 クラスの判別問題を考える。SVM では、 N 次元の法線ベクトル w および実数 b で定まる線形関数 $f(x) = x^T w - b$ を用い判別を行うことを考える。本論文では、各データ A_j を第 j 行ベクトルとする M 行 N 列の行列 A 、各ラベルの値を要素とする M 次元ベクトル y 、および M 次対角行列 Y をそれぞれ

$$A = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_M \end{bmatrix}, \quad y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{pmatrix}, \quad Y = \begin{bmatrix} y_1 & 0 & \cdots & 0 \\ 0 & y_2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & y_M \end{bmatrix}$$

と定義する。さらに、 e を要素が全て 1 のベクトル、また $\xi = (\xi_1, \xi_2, \dots, \xi_M)^T$ と定める。このとき、Vapnik[10] は次の最適化問題:

$$(2.1) \quad \begin{cases} \text{最小化} & 1/2 \|w\|^2 + Ce^T \xi \\ \text{制約} & \xi \geq e - YAw + yb, \quad \xi \geq 0 \end{cases}$$

を解くことで、 w, b を定める 1-norm SVM を提案している。ここで、 C は予め定められた正のパラメータである。

SVM が非常に広範囲の分野で高い判別力を発揮できるのは、カーネル関数を用いることで非線形な判別関数が構成できる点にある。非線形な判別関数を構成するために、まず適当な非線形写像

$\phi: \mathbb{R}^N \rightarrow \mathcal{F}$ を使い, 各データ A_j をより高い次元の空間 \mathcal{F} へと非線形に変換する. 非線形 SVM では, 変換された \mathcal{F} の点 $\phi(A_1), \phi(A_2), \dots, \phi(A_M)$ に対して問題 (2.1) を適用し, \mathcal{F} 上での線形判別関数を求めることで, 元の空間での非線形な判別関数を構成する. そこで, (2.1) の双対問題に対して最適化を試みる. 問題 (2.1) の場合, $\alpha \in \mathbb{R}^M$ を双対変数として, Wolfe [6] の双対問題を作れば, 以下の二次計画問題

$$(2.2) \quad \begin{cases} \text{最小化} & W(\alpha) = \frac{1}{2} \alpha^T Q \alpha - e^T \alpha \\ \text{制約} & y^T \alpha = 0, \quad 0 \leq \alpha \leq eC \end{cases}$$

が得られる. ただし Q は M 次の正方行列で, $i-j$ 要素 Q_{ij} がベクトル A_i と A_j の内積 $\langle A_i, A_j \rangle$ によって

$$(2.3) \quad Q_{ij} = y_i y_j \langle A_i, A_j \rangle$$

と定められる行列である. あるいは, 行列を使い表現すれば,

$$(2.4) \quad Q = Y A A^T Y$$

である. また, 双対問題の最適解と主問題の最適解をそれぞれ α^* および (w^*, b^*) とすれば, 両者の間には,

$$(2.5) \quad w^* = A^T Y \alpha^* = \sum_{j=1}^M y_j \alpha_j^* A_j^T$$

となる関係が, また, b^* は双対問題の制約式 $y^T \alpha = 0$ に対応した主問題の変数であり, 相補性条件より, $C > \alpha_j^* > 0$ となる任意の添え字 j に対し $b^* = y_j - A_j^T w^*$ となる関係が導かれる.

非線形変換されたデータの場合でも, 同様な方法で双対問題が構成可能である. 変換された \mathcal{F} の元 $\phi(A_i)$ と $\phi(A_j)$ の内積の値を $i-j$ 要素とする M 次正方行列を K とすれば, 変換された点に対する 1-norm SVM の双対問題は, 式 (2.4) の $A A^T$ の部分を K に置き換えたものに他ならない. すなわち, 双対問題を定めるためには行列 K を定義すればよく, 必ずしも特徴空間 \mathcal{F} の元 $\phi(A_j)$ を陽に表現する必要が無い.

そこで, SVM ではカーネル関数と呼ばれる特殊な関数を用い, $A_i, A_j \in \mathbb{R}^N$ から直接 K を算出し双対問題を定めている. さらに, \mathcal{F} での最適な法線ベクトルは, 式 (2.5) より

$$\sum_{j=1}^M y_j \alpha_j^* \phi(A_j) \in \mathcal{F}$$

となることから, 未知のデータ x を判別する場合には

$$\left\langle \sum_{j=1}^M y_j \alpha_j^* \phi(A_j), \phi(x) \right\rangle - b^* = \sum_{j=1}^M y_j \alpha_j^* \langle \phi(A_j), \phi(x) \rangle - b^* = \sum_{j=1}^M y_j \alpha_j^* K(A_j, x) - b^*$$

の正負を計算すればよく, これも, $\phi(x)$ を陽に求めることなく計算が行える. なお, カーネル関数には様々のものが知られており, $K(x, z) = (x^T z + 1)^d$ となるカーネル関数を polynomial kernel と呼び, これを用いれば, d 次以下の多項式の判別関数が構成可能である. その他にも SVM でよく用いられる代表的なカーネル関数として, RBF kernel $K(x, z) = \exp(-\|x - z\|^2 / \sigma^2)$, や sigmoid kernel $K(x, z) = \tanh(\kappa x^T z - \Theta)$ (ただし d は自然数のパラメータ, σ, κ, Θ は実数のパラメータである), などがあり, 様々の分野 [3, 7, 4] で用いられている.

3 SVM に対する最適化手法

この節では、特に問題 (2.1) で定式化される 1-norm SVM に対する最適化アルゴリズムについて考える。

一般には、主問題 (2.1) や双対問題 (2.2) は凸二次計画問題であることから、内点法 [9] などの多項式オーダーの最適化アルゴリズムを用い、容易に最適化可能であると考えられる。しかし、データマイニングに見られる多くの問題では、データの数 (M) が極めて大きな場合を扱わなくてはならない。例えば、カーネルを用いた双対問題 (2.2) では、 $M \times M$ の大型で完全に稠密な行列 $Q = YKY$ を扱わなくてはならない。 M が数万を超えるような状況では、特殊な計算機環境以外では Q を主記憶に配列として保持することは不可能であろう。よって、汎用的な最適化パッケージを用いて扱うことは困難である。そこで、双対問題の制約式が一本のみであることや、最適解では多くの変数が下限か上限に一致しているといった特徴を使い、 M が数万を超える大規模問題も扱える特殊な解法が提案されている。これらのアルゴリズムを実装した多くのソフトウェアパッケージも提供されており、 SVM^{light} [5], SVM^{Torch} [2] あるいは $LIBSVM$ [1] などがよく用いられているようである。

3.1 切除平面を用いた解法

この節では、主問題 (2.1) を切除平面法を用いて解くアルゴリズムについて述べる。まず、

$$(3.1) \quad \xi(w, b) \equiv \sum_{j=1}^M \max \{0, 1 - y_j (A_j w - b)\}$$

と定め、問題 (2.1) を

$$(3.2) \quad \left| \begin{array}{l} \text{最小化} \quad \psi(w, b) = 1/2 \|w\|^2 + C\xi(w, b) \end{array} \right.$$

と無制約の最適化問題として書き換える。その上で、問題 (2.1) の下界値を求めることを考えよう。

いま、任意の添え字の部分集合 $J \subseteq \{1, 2, \dots, M\}$ を使い、線形関数

$$(3.3) \quad \xi_J(w, b) = \sum_{j \in J} \{1 - y_j (A_j w - b)\}$$

を定める。 $\xi(w, b)$ は区分的に線形な凸関数であることより、 $\xi_J(w, b)$ は $\xi(w, b)$ を下から近似する線形関数である。添え字集合 J の代わりに、 M 次の 0-1 ベクトル $h = (h_1, h_2, \dots, h_M)^T$ を

$$h_j = \begin{cases} 1 & \text{if } j \in J, \\ 0 & \text{if } j \notin J \end{cases}$$

とすれば、

$$(3.4) \quad \xi_J(w, b) = \sum_{j \in J} \{1 - y_j (A_j w - b)\} = -h^T Y A w + h^T y b + h^T e$$

と記述できるから、任意に 0-1 ベクトル $h \in \{0, 1\}^M$ を定めることで、 $\xi(w, b)$ を下から近似する線形関数が定義される。すなわち、

$$\xi(w, b) \geq -h^T Y A w + h^T y b + h^T e, \quad \forall w, b$$

である。

次に、できるだけ大きな下界が得られるようベクトル h を定めることを考えよう。そのための準備として、問題 (2.1) を $w = w^0$ と固定したもとで変数 b に関してのみ最適化することを考える。(2.1) は、変数 w に関する目的関数の二次の項が定数になったことより、次の線形計画問題

$$(3.5) \quad \begin{cases} \text{最小化} & e^T \xi \\ \text{制約} & \xi \geq yb + \eta^0, \quad \xi \geq 0 \end{cases}$$

に帰着される。ここで、 η^0 は $\eta^0 = e - YAw^0$ と定められる M 次の定数ベクトルである。さらに、問題 (3.5) の双対問題は、 $\pi \in \mathbb{R}^M$ を双対変数として

$$(3.6) \quad \begin{cases} \text{最大化} & \pi^T \eta^0 \\ \text{制約} & y^T \pi = 0, \quad 0 \leq \pi \leq e \end{cases}$$

と記述される。

問題 (3.6) は、変数の上下限制約と等式制約が 1 つのみの連続ナップザック問題である。しかも $y \in \{-1, 1\}^M$ であることより 0-1 の最適解が存在する。そこで、 m をラベルが「1」であるデータの個数として、データの添え字に関して、

$$(3.7) \quad y_1 = y_2 = \cdots = y_m = +1, \quad y_{m+1} = y_{m+2} = \cdots = y_M = -1$$

を仮定する。また η^0 の各要素が

$$(3.8) \quad \eta_1^0 \geq \eta_2^0 \geq \cdots \geq \eta_m^0, \quad \eta_M^0 \geq \eta_{M-1}^0 \geq \cdots \geq \eta_{m+1}^0$$

とソートされていたと仮定する。もし、

$$\eta_{t^*}^0 + \eta_{M-t^*+1}^0 \geq 0 > \eta_{t^*+1}^0 + \eta_{M-t^*}^0$$

となる添え字 t^* が存在すれば、

$$\begin{aligned} \pi_1 = \pi_2 = \cdots = \pi_{t^*} = 1, \quad \pi_{t^*+1} = \pi_{t^*+2} = \cdots = \pi_m = 0, \\ \pi_M = \pi_{M-1} = \cdots = \pi_{M-t^*+1} = 1, \quad \pi_{M-t^*} = \pi_{M-t^*-1} = \cdots = \pi_{m+1} = 1 \end{aligned}$$

が双対問題 (3.6) の最適解である。

また、以下の定理が成り立つことが確かめられる。

定理 3.1. 任意に定めたベクトル $w^0 \in \mathbb{R}^N$ に対し、 b^0 を主問題 (3.5) の最適解、 h^0 を双対問題 (3.6) の 0-1 の最適解とする。このとき、不等式

$$\xi(w, b) \geq -h^{0T} YAw + h^{0T} e, \quad \forall w, b$$

が成立する。特に $(w, b) = (w^0, b^0)$ においては等号が成立する。

以上の議論に基づき、次のように問題 (2.1) の下界を求める。いま、 s 個のベクトル $w^k \in \mathbb{R}^N$ ($k = 1, 2, \dots, s$) および、それぞれに対応した問題 (3.6) の 0-1 の最適解 h^k が次のように与えられていると仮定する。

$$h^k = \operatorname{argmax} \{ \pi^T \eta^k \mid y^T \pi = 0, \quad 0 \leq \pi \leq e \}, \quad k = 1, 2, \dots, s.$$

ただし $\eta^k = e - YAw^k$ である. 0-1 ベクトル h^k に対応した s 個の線形関数 $-h^{kT}YAw + h^{kT}e$ ($k = 1, 2, \dots, s$) はいずれも $\xi(w, b)$ を下から線形近似することより, 次の凸二次計画問題:

$$(3.9) \quad \begin{cases} \text{最小化} & 1/2\|w\|^2 + Cr \\ \text{制約} & r \geq -h^{kT}YAw + h^{kT}e, \quad k = 1, 2, \dots, s \end{cases}$$

は, 元の問題 (2.1) の下界を与える.

問題 (3.9) の最適解を (w^{s+1}, r^{s+1}) とする. さらに, w^{s+1} に対して同様な方法で 0-1 ベクトル h^{s+1} を求めたとする. すなわち, $\eta^{s+1} = e - YAw^{s+1}$ として 0-1 ベクトル h^{s+1} を

$$(3.10) \quad h^{s+1} = \operatorname{argmax} \{ \pi^T \eta^{s+1} \mid y^T \pi = 0, \quad 0 \leq \pi \leq e \}$$

と求めたとする. このとき次の定理が成り立つ.

定理 3.2. 以下の等式

$$(3.11) \quad r^{s+1} = h^{s+1T} \eta^{s+1}$$

が成立すれば, w^{s+1} は問題 (2.1) の最適解である.

証明 問題 (2.1) の最適値を ψ^* と記す. 問題 (3.9) は常に下界を与えるのだから,

$$(3.12) \quad 1/2\|w^{s+1}\|^2 + Cr^{s+1} \leq \psi^*$$

である. 一方, 問題 (3.10) に対応する主問題

$$(3.13) \quad \begin{cases} \text{最小化} & e^T \xi \\ \text{制約} & \xi \geq yb + \eta^{s+1}, \quad \xi \geq 0 \end{cases}$$

の最適解を (ξ^{s+1}, b^{s+1}) とすれば, これは元問題 (2.1) の実行可能解であることより, $\psi(\xi^{s+1}, b^{s+1})$ は ψ^* の上界となる. また, 双対定理より $e^T \xi^{s+1} = h^{s+1T} \eta^{s+1}$ なので,

$$(3.14) \quad \psi^* \leq 1/2\|w^{s+1}\|^2 + Ce^T \xi^{s+1} = 1/2\|w^{s+1}\|^2 + Ch^{s+1T} \eta^{s+1}$$

である. そこで, (3.12) と (3.14) より

$$1/2\|w^{s+1}\|^2 + Cr^{s+1} \leq \psi^* \leq 1/2\|w^{s+1}\|^2 + Ch^{s+1T} \eta^{s+1}$$

を得る. ゆえに, $r^{s+1} = h^{s+1T} \eta^{s+1}$ ならば $(w^{s+1}, \xi^{s+1}, b^{s+1})$ は最適解となる. \square

一方,

$$r^{s+1} < h^{s+1T} \eta^{s+1}$$

であった場合には, 0-1 ベクトル h^{s+1} で定まる新たな $s+1$ 番目の不等式

$$r \geq -h^{s+1T}YAw + h^{s+1T}e$$

を問題 (3.9) に追加することで, 解 (w^{s+1}, r^{s+1}) を切除することが可能である.

したがって, 終了判定に用いる精度 $\epsilon > 0$ を定めれば, 図 1 に示した切除平面法によるアルゴリズムが構成できる. また, 追加できる不等式の数是有限であることより, 以下の定理が成立する.

任意の初期解 w^1 を定め, $\eta^1 = e - YAw^1$ として 0-1 ベクトル

$$h^1 = \pi^* = \operatorname{argmax} \{ \pi^T \eta^1 \mid y^T \pi = 0, 0 \leq \pi \leq e \}$$

を求める. カウンタを $s = 1$ に初期化する.

ステップ 1. 二次計画問題 (3.9) を解き, その最適解を (w^{s+1}, r^{s+1}) とする.

ステップ 2. $\eta^{s+1} = e - YAw^{s+1}$ として 0-1 ベクトル h^{s+1} を

$$h^{s+1} = \pi^* = \operatorname{argmax} \{ \pi^T \eta^{s+1} \mid y^T \pi = 0, 0 \leq \pi \leq e \}$$

と求める.

$$\frac{h^{s+1T} \eta^{s+1} - r^{s+1}}{|r^{s+1}|} \leq \epsilon \text{ ならば } w^{s+1} \text{ を解として終了する.}$$

ステップ 3. カウンタを $s = s + 1$ と更新してステップ 1. へ戻る.

図 1: アルゴリズム Prim

定理 3.3. 有限回の不等式の追加で, 元問題 (2.1) の最適解が生成される.

次に, 上で述べたアルゴリズムを, カーネルを使った非線形判別を行う問題へ拡張しよう. 前節で述べたように, AA^T となる箇所をカーネル行列 K に置き換えを行えばよい.

まず, 準備として, s 個の 0-1 ベクトル h^k ($k = 1, 2, \dots, s$) を列ベクトルとした $M \times s$ 行列

$$H^s = [h^1 \ h^2 \ \dots \ h^s]$$

を定義する. 行列 H^s を使えば, 問題 (3.9) は

$$(3.15) \quad \begin{cases} \text{最小化} & 1/2 \|w\|^2 + Cr \\ \text{制約} & er \geq -H^{sT} Y A w + H^{sT} e \end{cases}$$

と書き換えられる. そこで, 制約条件の各行に対応した s 次の変数ベクトル $\beta \in \mathbb{R}^s$ を用いて双対問題作れば,

$$(3.16) \quad \begin{cases} \text{最小化} & 1/2 \beta^T H^{sT} Y A A^T Y H^s \beta - e^T H^s \beta \\ \text{制約} & e^T \beta = C, \beta \geq 0 \end{cases}$$

を得る. この問題は AA^T の部分をカーネル行列 K に置き換えることで, 非線形判別問題に対応可能である.

また, 問題 (3.16) の最適解を $\beta^{s+1} \in \mathbb{R}^s$, また対応する主問題 (3.15) の最適解を (w^{s+1}, r^{s+1}) とすれば, これらの間には次の関係が成り立つ. まず,

$$w^{s+1} = A^T Y H^s \beta^{s+1}$$

であることを使い,

$$\eta^{s+1} = e - YAw^{s+1} = e - YAA^T YH^s \beta^{s+1}$$

と変形できる. やはり AA^T の部分をカーネル行列 K に置き換えれば, ベクトル η^{s+1} が計算可能である. よって,

$$r^{s+1} = \max_{k=1,2,\dots,s} h^k{}^T \eta^{s+1},$$

を得ることができる. なお, r^{s+1} は問題 (3.16) の等式制約に対応した最適な双対変数として算出することも可能である.

さらに, $\alpha^{s+1} \equiv H^s \beta^{s+1}$ とすれば,

$$y^T \alpha^{s+1} = 0, \quad 0 \leq \alpha^{s+1} \leq eC$$

であることより, $W(\alpha^k)$ により上界値も得ることが可能である. 図 2 にカーネル関数を用いた場合のアルゴリズムを示した.

任意にベクトル $\alpha^1 \in \mathbb{R}^M$ を定め, $\eta^1 = e - YKY\alpha^1$ とする. 0-1 ベクトル

$$h^1 = \pi^* = \operatorname{argmax} \{ \pi^T \eta^1 \mid y^T \pi = 0, \quad 0 \leq \pi \leq e \}$$

を求め $H^1 = [h^1]$ とする. カウンタを $s = 1$ に初期化する.

ステップ 1. 以下の二次計画問題

$$(3.17) \quad \begin{cases} \text{最小化} & 1/2 \beta^T H^s{}^T YKYH^s \beta - e^T H^s \beta \\ \text{制約} & e^T \beta = C, \quad \beta \geq 0. \end{cases}$$

を解き, 最適解を β^{s+1} とする. また, 等式制約に対応する最適な双対変数を r^{s+1} とする.

ステップ 2. $\eta^{s+1} = e - YKYH^s \beta^{s+1}$ として 0-1 ベクトル h^{s+1} を

$$(3.18) \quad h^{s+1} = \pi^* = \operatorname{argmax} \{ \pi^T \eta^{s+1} \mid y^T \pi = 0, \quad 0 \leq \pi \leq e \}$$

と求める.

$$\frac{h^{s+1}{}^T \eta^{s+1} - r^{s+1}}{|r^{s+1}|} \leq \epsilon \text{ ならば } \alpha^* = H^s \beta^{s+1} \text{ を最適解として終了する.}$$

ステップ 3. $H^{s+1} = [h^1 \dots h^s h^{s+1}]$ とする. カウンタを $s = s + 1$ と更新してステップ 1. へ戻る.

図 2: アルゴリズム Kernel

図 2 のアルゴリズムで最も計算を必要とする部分は, ステップ 1. における問題 (3.17) の目的関数係数行列

$$P^s = H^s{}^T YKYH^s$$

の構成, およびステップ 2. での η^{s+1} の計算に必要な

$$YKYH^s\beta^{s+1}$$

の計算である. どちらの部分も K が稠密で非常に大きな行列となり, 主記憶に保持できない場合には実装上の工夫が不可欠である.

まず, ベクトル

$$g^k = YKYh^k, \quad k = 1, 2, \dots, s$$

を定めて, これらを列ベクトルとした $M \times s$ 行列を

$$G^s = [g^1 \ g^2 \ \dots \ g^s] = YKYH^s$$

とする. このとき

$$P^s = H^{sT}YKYH^s = H^{sT}G^s$$

となる関係が得られる. また行列 G^s を使えば,

$$\eta^{s+1} = e - YKYH\beta^{s+1} = e - G^s\beta^{s+1}$$

と計算されることから, $M \times s$ 行列 H^s および G^s を配列として主記憶に保持することで, 計算の効率化が図られる. さらに, P^s も保持されているとすれば, $s+1$ 回目の繰り返しのステップ 1. では

$$P^{s+1} = H^{s+1T}G^{s+1} = \begin{bmatrix} P^s & H^{sT}g^{s+1} \\ h^{s+1T}G^s & h^{s+1T}g^{s+1} \end{bmatrix}$$

となることより, 新たに必要な計算は, $H^{sT}g^{s+1}$ および $h^{s+1T}g^{s+1}$ となる.

以上の手順で計算する場合, 最も計算量が必要な部分はベクトル $g^k = YKYh^k$ の計算である. h^k が 0-1 ベクトルであることから, h^k の要素が 1 となる添え字に対応した行列 K の列ベクトルの算出が必要となる.

アルゴリズムの初期段階の繰り返しにおいて生成される h^k は, 多くの 1 を要素に持つことが予想され, その結果, 行列 K の多くの列ベクトルの計算が必要となり計算時間の増加につながってしまう. そこで, アルゴリズムのステップ 2. において問題 (3.18) の最大化は行わず,

$$(3.19) \quad h^T\eta^{s+1} > r^{s+1}, \quad y^Th = 0, \quad h \in \{0, 1\}^M$$

となる 0-1 ベクトル h を用いることにする. h がこれを満たせば, 不等式

$$r \geq h^TYAw + h^Te$$

によって (w^{s+1}, r^{s+1}) は切除され, 下界値が更新されることが期待できる. そこで, (3.19) を満たす h で, できるだけ 0 の要素を多く持つものをヒューリスティックを使い求めることで, g^{s+1} の計算の高速化が期待できる.

3.2 ν -サポートベクターアルゴリズムへの拡張

近年, Schölkopf 等 [8] は問題 (2.1) に新たな変数 ρ を導入して, 次の最適化問題

$$(3.20) \quad \begin{cases} \text{最小化} & \frac{1}{2}\|w\|^2 - \nu\rho + \frac{1}{M}e^T\xi \\ \text{制約} & \xi \geq e\rho - YAw + yb, \quad \xi \geq 0, \quad \rho \geq 0 \end{cases}$$

を解き w および b を算出する ν -SVC を提案している. ただし, ν は問題 (2.1) の C に相当するパラメータで, $0 \leq \nu \leq 1$ である. まお, 2.1 節の 1-norm SVM の定式化 (2.1) の場合と同様, $Q = YAA^TY$ と定め双対問題を作れば,

$$(3.21) \quad \begin{cases} \text{最小化} & \frac{1}{2}\alpha^T Q \alpha \\ \text{制約} & y^T \alpha = 0, \quad e^T \alpha \geq \nu, \quad 0 \leq \alpha \leq \frac{1}{M} \end{cases}$$

となる.

前節で述べた切除平面法のアルゴリズムは, 以下のようにして適用が可能である. まず,

$$(3.22) \quad \xi^\nu(w, b, \rho) \equiv \sum_{j=1}^M \max \{0, \rho - y_j A_j w + y_j b\} - \nu M \rho$$

と誤差関数を定め, (3.20) を

$$(3.23) \quad \begin{cases} \text{最小化} & \frac{1}{2}\|w\|^2 + \frac{1}{M}\xi^\nu(w, b, \rho) \\ \text{制約} & \rho \geq 0 \end{cases}$$

のように区分的に線形な凸関数の最小化問題として扱うことにする.

次に関数 $\xi^\nu(w, b, \rho)$ を下から近似する線形関数を求め, (3.23) の下界値を求めよう. これには, $w = w^0$ と固定したもとので, 問題 (3.20) を解けばよい. 定数部分を

$$\eta^0 \equiv -YAw^0$$

とすれば, (3.20) は

$$(3.24) \quad \begin{cases} \text{最小化} & -\nu M \rho + e^T \xi \\ \text{制約} & \xi - e\rho - yb \geq \eta^0, \quad \xi \geq 0, \quad \rho \geq 0 \end{cases}$$

と線形計画問題に帰着される. なお, 目的関数は M 倍して定数項は削除した. この双対問題は $\pi \in \mathbb{R}^M$ を双対変数とすれば,

$$(3.25) \quad \begin{cases} \text{最大化} & \pi^T \eta^0 \\ \text{制約} & y^T \pi = 0, \quad e^T \pi \geq \nu M, \quad 0 \leq \pi \leq e \end{cases}$$

と書き表される. これは, 問題 (3.6) に制約 $e^T \pi \geq \nu M$ が付け加えられたものである. したがって, (3.6) の最適解 π^* がもし $e^T \pi^* \geq \nu M$ を満たせば, (3.25) の最適化となる. $e^T \pi^* < \nu M$ の場合, 前節と同様に, (3.7) および (3.8) を仮定すれば,

$$t^* = \left\lfloor \frac{\nu M}{2} \right\rfloor$$

となる添え字 t^* を定め,

$$\pi_1 = \pi_2 = \cdots = \pi_{t^*} = 1, \quad \pi_{t^*+1} = \frac{\nu M}{2} - t^*, \quad \pi_{t^*+2} = \pi_{t^*+3} = \cdots = \pi_m = 0,$$

$$\pi_M = \pi_{M-1} = \cdots = \pi_{M-t^*+1} = 1, \quad \pi_{M-t^*} = \frac{\nu M}{2} - t^*, \quad \pi_{M-t^*-1} = \pi_{M-t^*-2} = \cdots = \pi_{m+1} = 0,$$

としたものが, 双対問題 (3.25) の最適解である.

このとき, 定理 3.1 と同様な, 以下の定理が成立する.

定理 3.4. 任意のベクトル $w^0 \in \mathbb{R}^n$ に対応して定まる問題 (3.25) の最適解を h^0 とすれば, 不等式

$$(3.26) \quad \xi^\nu(w, b, \rho) \geq (e^T h^0 - \nu M) \rho - h^{0T} Y A w, \quad \forall w, b, \rho$$

が成立する. また, 問題 (3.24) の最適解を b^0, ρ^0 とすれば, $\rho^0 = 0$ となるものが存在し, 不等式は $(w, b, \rho) = (w^0, b^0, 0)$ のとき等号が成立している.

いま, s 個のベクトル h^k ($k = 1, 2, \dots, s$) が得られたと仮定すれば, 次の二次計画問題:

$$(3.27) \quad \left\{ \begin{array}{ll} \text{最小化} & \frac{1}{2} \|w\|^2 + \frac{1}{M} r \\ \text{制約} & r \geq (e^T h^k - \nu M) \rho - h^{kT} Y A w, \quad k = 1, 2, \dots, s \\ & \rho \geq 0 \end{array} \right.$$

が得られる. $e^T h^k - \nu M$ が非負であることから, $\rho = 0$ となる最適解が存在する. したがって, この問題は

$$(3.28) \quad \left\{ \begin{array}{ll} \text{最小化} & \frac{1}{2} \|w\|^2 + \frac{1}{M} r \\ \text{制約} & r \geq -h^{kT} Y A w, \quad k = 1, 2, \dots, s \end{array} \right.$$

と簡単に書け, これを解くことで (3.20) の下界が得られる. また, (3.28) の最適解を (w^{s+1}, r^{s+1}) として,

$$(3.29) \quad \eta^{s+1} = -Y A w^{s+1}$$

を求め, 新たな 0-1 ベクトル h^{s+1} を

$$(3.30) \quad h^{s+1} = \operatorname{argmax} \{ \pi^T \eta^{s+1} \mid y^T \pi = 0, \quad e^T \pi \geq \nu M, \quad 0 \leq \pi \leq e \}$$

と求めたとする. このとき定理 3.2 と同様に

$$r^{s+1} = h^{s+1T} \eta^{s+1}$$

が成立すれば, w^{s+1} は (3.20) の最適解であることが証明できる.

また,

$$H^s = [h^1 \ h^2 \ \cdots \ h^s]$$

と定義して、制約条件の各行に対応した s 次の変数ベクトル $\beta \in \mathbb{R}^s$ を用いて (3.28) の双対問題作れば、

$$(3.31) \quad \begin{cases} \text{最小化} & \frac{1}{2} \beta^T H^s Y A A^T Y H^s \beta \\ \text{制約} & e^T \beta = \frac{1}{M}, \quad \beta \geq 0 \end{cases}$$

となることから、 AA^T 部分を \mathcal{K} に置き換えることで、カーネルを用いた非線形な判別問題の場合には、

$$(3.32) \quad \begin{cases} \text{最小化} & \frac{1}{2} \beta^T H^s Y \mathcal{K} Y H^s \beta \\ \text{制約} & e^T \beta = \frac{1}{M}, \quad \beta \geq 0 \end{cases}$$

を解けばよい。さらに、問題 (3.31) の最適解を β^{s+1} 、対応した主問題 (3.27) の最適解を $(w, b, r, \rho) = (w^{s+1}, b^{s+1}, r^{s+1}, 0)$ とすれば、

$$w^{s+1} = A^T Y H^s \beta^{s+1}$$

となる関係を使えば、

$$\eta^{s+1} = -Y A w^{s+1} = -Y A A^T H^{s+1} \beta^{s+1}$$

となり、 AA^T の部分を \mathcal{K} にすれば、カーネル関数を用いた場合も、問題 (3.25) を定めることが可能である。また、 $\alpha^{s+1} = H^s \beta^{s+1}$ とすれば、 α^{s+1} は

$$y^T \alpha^{s+1} = 0, \quad e^T \alpha^{s+1} \geq \nu, \quad 0 \leq \alpha^{s+1} \leq \frac{1}{M}$$

となり、双対問題 (3.21) の制約式を満たしている。

4 サポートベクター回帰

この節では、上で導入した切除平面法による枠組みを拡張し、サポートベクター回帰へと適応する。すなわち、各データ $j = 1, 2, \dots, M$ において、 N 次の属性ベクトル $A_j \in \mathbb{R}^N$ および実数 y_j が与えられたもとで、線形関数 $f(x) = x^T w - b$ を用い y_j を回帰する問題を考える。

標準的に用いられているサポートベクター回帰の枠組みでは、ある正の ϵ を定め、以下の式で与えられる ϵ を許容した誤差項

$$\xi_j = \max \{ 0, |y_j - (A_j w - b)| - \epsilon \}$$

の総和をできるだけ小さくするよう w および b を求めることを考える。その際、1-norm SVM の場合と同様に、 $\|w\|$ の最小化も合わせて考え、以下の最適化問題:

$$(4.1) \quad \begin{cases} \text{最小化} & \frac{1}{2} \|w\|^2 + \frac{C}{M} \sum_{j=1}^M \xi_j \\ \text{制約} & \xi_j \geq |y_j - A_j w + b| - \epsilon, \quad \xi_j \geq 0, \quad j = 1, 2, \dots, M \end{cases}$$

任意にベクトル $\alpha^1 \in \mathbb{R}^M$ を定め, $\eta^1 = -YKY\alpha^1$ とする. 0-1 ベクトル

$$h^1 = \pi^* = \operatorname{argmax} \{ \pi^T \eta^1 \mid y^T \pi = 0, \quad e^T \pi \geq \nu M, \quad 0 \leq \pi \leq e \}$$

を求め $H^1 = [h^1]$ とする. カウンタを $s = 1$ に初期化する.

ステップ 1. 以下の二次計画問題

$$(3.33) \quad \begin{cases} \text{最小化} & 1/2 \beta^T H^s T Y K Y H^s \beta \\ \text{制約} & e^T \beta = \frac{1}{M}, \quad \beta \geq 0 \end{cases}$$

を解き, 最適解を β^{s+1} とする. また, 等式制約に対応する最適な双対変数を r^{s+1} とする.

ステップ 2. $\eta^{s+1} = -YKYH^s \beta^{s+1}$ として 0-1 ベクトル h^{s+1} を

$$(3.34) \quad h^{s+1} = \pi^* = \operatorname{argmax} \{ \pi^T \eta^{s+1} \mid y^T \pi = 0, \quad e^T \pi \geq \nu M, \quad 0 \leq \pi \leq e \}$$

と求める. $\frac{h^{s+1T} \eta^{s+1} - r^{s+1}}{|r^{s+1}|} \leq \epsilon$ ならば $\alpha^* = H^s \beta^{s+1}$ を最適解として終了する.

ステップ 3. $H^{s+1} = [h^1 \dots h^s h^{s+1}]$ とする. カウンタを $s = s + 1$ と更新してステップ 1. へ戻る.

図 3: アルゴリズム **Kernel** for ν -SVC

を解き, w および b を求める. ただし, C は予め定められた正のパラメータである. これは, Vapnik [10] によるサポートベクター回帰の定式化でこれ以降 ϵ -SVR と呼ぶことにする.

一方, 近年 Schölkopf 等 [8] は次の最適化問題:

$$(4.2) \quad \begin{cases} \text{最小化} & \frac{1}{2} \|w\|^2 + \frac{C}{M} \left(\nu M \epsilon + \sum_{j=1}^M \xi_j \right) \\ \text{制約} & \xi_j \geq |y_j - A_j w + b| - \epsilon, \quad \xi_j \geq 0, \quad j = 1, 2, \dots, M, \\ & \epsilon \geq 0 \end{cases}$$

を解き, w および b を算出する ν -SVR を提案している. ただし, この問題の場合には, ϵ は最適化される変数であり, ν および C はある与えられた正のパラメータである.

SVM の場合と同様, どちらの定式化も, 双対問題を考えることでカーネル関数を用いた非線形な回帰が行える. α を M 次の双対変数ベクトル, また $K = AA^T$ と定めれば, 問題 (4.1) および問題 (4.2) の双対問題は, それぞれ,

$$(4.3) \quad \begin{cases} \text{最小化} & \frac{1}{2} \alpha^T K \alpha + \epsilon e^T |\alpha| - y^T \alpha \\ \text{制約} & e^T \alpha = 0, \quad -e \frac{C}{M} \leq \alpha \leq e \frac{C}{M} \end{cases}$$

および

$$(4.4) \quad \begin{cases} \text{最小化} & \frac{1}{2} \alpha^T K \alpha - y^T \alpha \\ \text{制約} & e^T \alpha = 0, \quad e^T |\alpha| \leq \nu C, \quad -e \frac{C}{M} \leq \alpha \leq e \frac{C}{M} \end{cases}$$

と表される。ただし、 $|\alpha|$ は α の各要素の絶対値からなる M 次ベクトルと約束する。ここで、行列 K の $i-j$ 要素を、カーネル関数 $K(A_i, A_j)$ と定めれば、非線形な回帰が行えることとなる。 α^* を双対問題の最適解とすれば、求めたい回帰式は

$$f(x) = \sum_{j=1}^M \alpha_j K(A_j, x) - b^*$$

となる。なお、 b^* は主問題 (4.1) あるいは (4.2) の変数 b の最適解であり、双対問題の制約条件 $e^T \alpha = 0$ に対応した変数である。

以降の節では、この2つの定式化に対して、切除平面法によるアプローチを行う。

4.1 ϵ -SVR による回帰

まず、 ϵ -SVR に対し切除平面法による解法を示す。ある正の ϵ を許容した誤差項を w と b の関数として、

$$L^\epsilon(w, b) \equiv \sum_{j=1}^M \max \{ 0, |y_j - A_j w + b| - \epsilon \}$$

と定義し、問題 (4.1) を無制約の最適化問題:

$$(4.6) \quad \begin{cases} \text{最小化} & \frac{1}{2} \|w\|^2 + \frac{C}{M} L^\epsilon(w, b) \end{cases}$$

と考える。関数 $L^\epsilon(w, b)$ は区分的に線形な凸関数であることより、任意の点 $(w, b) = (w^0, b^0)$ において、添え字の部分集合 $J_+, J_- \subseteq \{1, 2, \dots, M\}$ を

$$J_+ = \{ j \mid y_j - A_j w^0 + b^0 - \epsilon \geq 0 \}, \quad J_- = \{ j \mid -y_j + A_j w^0 - b^0 - \epsilon \geq 0 \}$$

と定めれば、次の線形関数

$$(4.7) \quad L_{J_+, J_-}^\epsilon(w, b) \equiv \sum_{j \in J_+} (y_j - A_j w + b - \epsilon) + \sum_{j \in J_-} (-y_j + A_j w - b - \epsilon)$$

は、 $L^\epsilon(w, b)$ を下から近似する線形関数となる。添え字集合 J_+, J_- の代わりに、 M 次ベクトル $h = (h_1, h_2, \dots, h_M)^T$ を

$$h_j = \begin{cases} 1 & \text{if } j \in J_+, \\ -1 & \text{if } j \in J_-, \\ 0 & \text{o.w.} \end{cases}$$

とすれば、線形関数 (4.7) は

$$L_{J_+, J_-}^\epsilon(w, b) = h^T y - h^T A w + b e^T h - \epsilon e^T |h|$$

と記述される。ただし、 $|h| = (|h_1|, |h_2|, \dots, |h_M|)^T$ と約束する。

一般的に、以下の補題が成立している。

任意にベクトル $\alpha^1 \in \mathbb{R}^M$ を定め, $\eta^1 = y - K\alpha^1$ とする. ベクトル

$$h^1 = \pi^* = \operatorname{argmax} \{ \pi^T \eta^1 - \epsilon e^T |\pi| \mid e^T \pi = 0, \quad -e \leq \pi \leq e \}$$

および $\delta_1 = h^{1T} y - \epsilon e^T |h^1|$ を求め $H^1 = [h^1]$, $\delta^1 = (\delta_1)$ とする. カウンタを $s = 1$ に初期化する.

ステップ 1. 以下の二次計画問題

$$(4.5) \quad \left\{ \begin{array}{ll} \text{最小化} & 1/2 \beta^T H^s Y K Y H^s \beta - \delta^{sT} \beta \\ \text{制約} & e^T \beta = \frac{C}{M}, \quad \beta \geq 0 \end{array} \right.$$

を解き, 最適解を β^{s+1} とする. また, 等式制約に対応する最適な双対変数を r^{s+1} とする.

ステップ 2. $\eta^{s+1} = y - K H^s \beta^{s+1}$ として, ベクトル h^{s+1} および δ_{s+1} を

$$h^{s+1} = \pi^* = \operatorname{argmax} \{ \pi^T \eta^{s+1} - \epsilon e^T |\pi| \mid e^T \pi = 0, \quad -e \leq \pi \leq e \},$$

$$\delta_{s+1} = h^{s+1T} y - \epsilon e^T |h^{s+1}|$$

と求める. $\frac{h^{s+1T} \eta^{s+1} - \epsilon e^T |h^{s+1}| - r^{s+1}}{|r^{s+1}|} \leq \epsilon$ ならば $\alpha^* = H^s \beta^{s+1}$ を最適解として終了する.

ステップ 3. $H^{s+1} = [h^1 \dots h^s h^{s+1}]$, $\delta^{s+1} = (\delta_1 \dots \delta_s \delta_{s+1})$ とする. カウンタを $s = s + 1$ と更新してステップ 1. へ戻る.

図 4: アルゴリズム **Kernel** for ϵ -SVR

補題 4.1. h を $-e \leq h \leq e$ を満たす任意の M 次ベクトルとする. このとき以下の不等式が成り立つ.

$$(4.8) \quad L^\epsilon(w, b) \geq h^T y - h^T A w + b e^T h - \epsilon e^T |h|, \quad \forall w \in \mathbb{R}^N, b \in \mathbb{R}.$$

元問題 (4.1) において $w = w^0$ と固定したもとで b に関して最適化を行うことで, より強い下界値を与える線形関数を得ることが可能である. いま, 定数部分をまとめて, $\eta_j^0 = y_j - A_j w^0$ とすれば, (4.1) は以下の線形計画問題:

$$(4.9) \quad \left\{ \begin{array}{ll} \text{最小化} & \sum_{j=1}^M \xi_j \\ \text{制約} & \xi_j - b \geq \eta_j^0 - \epsilon, \quad \xi_j + b \geq -\eta_j^0 - \epsilon, \quad \xi_j \geq 0, \quad j = 1, 2, \dots, M \end{array} \right.$$

に帰着される. なお, 目的関数は C/M で除して定数項は省略した.

この線形計画問題は, 以下のように簡単に最適化が可能である. まず, (4.9) の双対問題は, $p, q \in$

\mathbb{R}^M を双対変数として

$$(4.10) \quad \left\{ \begin{array}{l} \text{最大化} \quad \sum_{j=1}^M \eta_j^0 (p_j - q_j) - \epsilon (p_j + q_j) \\ \text{制約} \quad \sum_{j=1}^M (p_j - q_j) = 0, \quad p_j + q_j \leq 1, \quad p_j, q_j \geq 0, \quad j = 1, 2, \dots, M \end{array} \right.$$

と表される. この問題の最適解に関して, 明らかに以下の性質が成り立っている.

補題 4.2. p^*, q^* を (4.10) の最適解とする. このとき, 全ての $j = 1, 2, \dots, M$ で, $p_j^* q_j^* = 0$ となる 0-1 の最適解が存在する. すなわち, p_j^*, q_j^* のどちらか一方は必ず 0 となる.

そこで, $\pi_j \equiv p_j - q_j$ と定めれば, 最適解では $|\pi_j| = p_j + q_j$ となることより, $|\pi| = (|\pi_1|, |\pi_2|, \dots, |\pi_M|)^T$ と記せば (4.10) は

$$(4.11) \quad \left\{ \begin{array}{l} \text{最大化} \quad \pi^T \eta^0 - \epsilon e^T |\pi| \\ \text{制約} \quad e^T \pi = 0, \quad -e \leq \pi \leq e \end{array} \right.$$

と書き直すことができる.

この問題の最適解は, 以下のように陽に求めることが可能である. なお, 簡単のため以降では η^0 の各要素に対して,

$$\eta_1^0 \geq \eta_2^0 \geq \dots \geq \eta_M^0$$

と仮定し議論を進めることとする.

いま, 以下の不等式

$$\eta_{t^*}^0 - \eta_{M-t^*+1}^0 - 2\epsilon \geq 0 > \eta_{t^*+1}^0 - \eta_{M-t^*}^0 - 2\epsilon$$

を満たす添え字 t^* を定め, M 次の 0-1 ベクトル $h^0 = (h_1^0, h_2^0, \dots, h_M^0)^T$ を

$$(4.12) \quad \begin{aligned} h_1^0 &= h_2^0 = \dots = h_{t^*}^0 = 1, \\ h_{t^*+1}^0 &= h_{t^*+2}^0 = \dots = h_{M-t^*}^0 = 0, \\ h_M^0 &= h_{M-1}^0 = \dots = h_{M-t^*+1}^0 = -1 \end{aligned}$$

とする.

定理 4.3. h^0 を (4.12) で定められた M 次の 0-1 ベクトルとする. このとき $\pi = h^0$ は問題 (4.11) の最適解である.

証明 添え字 t^* の定義より,

$$(4.13) \quad \min\{\eta_{t^*}^0 - \epsilon, \eta_{M-t^*}^0 + \epsilon\} \geq -b^0 \geq \max\{\eta_{t^*+1}^0 - \epsilon, \eta_{M-t^*+1}^0 + \epsilon\}$$

を満たす実数 b^0 が存在する. そこで, M 次ベクトル ξ^0 を

$$\begin{aligned} \xi_j^0 &= \eta_j^0 + b^0 - \epsilon, \quad j = 1, 2, \dots, t^*, \\ \xi_j^0 &= 0, \quad j = t^* + 1, t^* + 2, \dots, M - t^*, \\ \xi_j^0 &= -\eta_j^0 - b^0 - \epsilon, \quad j = M - t^* + 1, \dots, M - 1, M \end{aligned}$$

と定める。 b^0 が (4.13) を満たすことより, ξ^0 は主問題 (4.9) の実行可能解である。また, 簡単な計算をすれば

$$e^T \xi^0 = h^{0T} \eta^0 - \epsilon e^T |h^0|$$

となり, 主問題と双対問題の目的関数が一致することが分かる。したがって, h^0 は問題 (4.11) の最適解である。 \square

直ちに以下の補題が導かれる。

補題 4.4. 任意のベクトル $w^0 \in \mathbb{R}^n$ に対応し, $\eta^0 = y - Aw^0$ で定まる問題 (4.11) の最適解を h^0 とする。このとき, 不等式

$$L^\epsilon(w, b) \geq h^{0T} y - \epsilon e^T |h^0| - h^{0T} Aw, \quad \forall w, b$$

が成立する。また, $(w, b) = (w^0, b^0)$ において等号が成立する。

以上の結果を用いて, 前節で述べたものと全く同様な切除平面法のアルゴリズムが構成可能である。すなわち, 一般的にアルゴリズムの s 回目の繰り返しにおいて, s 個の 0-1 ベクトル $h^k \in \mathbb{R}^M$ ($k = 1, 2, \dots, s$) が得られたと仮定する。補題 4.4 より,

$$(4.14) \quad \left\{ \begin{array}{l} \text{最小化} \quad \frac{1}{2} \|w\|^2 + \frac{C}{M_T} r \\ \text{制約} \quad r \geq \delta_k - h^{kT} Aw, \quad k = 1, 2, \dots, s \end{array} \right.$$

を解くことで問題 (4.1) の下界を得ることができる。また,

$$H^s = [h^1 \ h^2 \ \dots \ h^s] \quad \text{および} \quad \delta^s = (\delta_1 \ \delta_2 \ \dots \ \delta_s)^T$$

と定義すれば, 双対問題は

$$(4.15) \quad \left\{ \begin{array}{l} \text{最小化} \quad \frac{1}{2} \beta^T H^{sT} A A^T H^s \beta - \delta^{sT} \beta \\ \text{制約} \quad e^T \beta = \frac{C}{M}, \quad \beta \geq 0 \end{array} \right.$$

となる。 β^* を双対問題の最適解とすれば, 主問題の最適解 w^* は

$$w^* = A^T H^s \beta^*$$

と求められる。このとき,

$$\alpha^* = H^s \beta^*$$

とすれば, α^* は問題 (4.3) の実行可能解となっている。したがって, カーネルを用いた非線形な回帰を行う場合には, AA^T の部分をカーネル行列 K に置き換え,

$$(4.16) \quad \left\{ \begin{array}{l} \text{最小化} \quad \frac{1}{2} \beta^T H^{sT} K H^s \beta - \delta^{sT} \beta \\ \text{制約} \quad e^T \beta = \frac{C}{M}, \quad \beta \geq 0 \end{array} \right.$$

を解けばよい。カーネルを用いた場合のアルゴリズムをまとめると図 4 のようになる。

任意にベクトル $\alpha^1 \in \mathbb{R}^M$ を定め, $\eta^1 = y - K\alpha^1$ とする. ベクトル

$$h^1 = \pi^* = \operatorname{argmax} \{ \pi^T \eta^1 \mid e^T \pi = 0, \quad -e \leq \pi \leq e, \quad e^T |\pi| \leq \nu M \}$$

を求め $H^1 = [h^1]$, とする. カウンタを $s = 1$ に初期化する.

ステップ 1. 以下の二次計画問題

$$(4.17) \quad \left\{ \begin{array}{ll} \text{最小化} & 1/2 \beta^T H^s Y K Y H^s \beta - y^T H^s \beta \\ \text{制約} & e^T \beta = \frac{C}{M}, \quad \beta \geq 0 \end{array} \right.$$

を解き, 最適解を β^{s+1} とする. また, 等式制約に対応する最適な双対変数を r^{s+1} とする.

ステップ 2. $\eta^{s+1} = y - K H^s \beta^{s+1}$ として, ベクトル h^{s+1} および δ_{s+1} を

$$h^{s+1} = \pi^* = \operatorname{argmax} \{ \pi^T \eta^{s+1} \mid e^T \pi = 0, \quad -e \leq \pi \leq e, \quad e^T |\pi| \leq \nu M \},$$

と求める. $\frac{h^{s+1 T} \eta^{s+1} - r^{s+1}}{|r^{s+1}|} \leq \epsilon$ ならば $\alpha^* = H^s \beta^{s+1}$ を最適解として終了する.

ステップ 3. $H^{s+1} = [h^1 \dots h^s h^{s+1}]$, とする. カウンタを $s = s + 1$ と更新してステップ 1. へ戻る.

図 5: アルゴリズム Kernel for ν -SVR

4.2 ν -SVR による定式化

Schölkopf 等 [8] による ν -SVR(4.2) に対しても, ほぼ同様な切除平面法のアプローチが可能である. まず, 誤差項に対応した関数を

$$L^\nu(w, b, \epsilon) \equiv \sum_{j=1}^M \max \{ 0, |y_j - A_j w + b| - \epsilon \} + \nu M \epsilon$$

と定め, 問題 (4.2) を

$$(4.18) \quad \left\{ \begin{array}{ll} \text{最小化} & \frac{1}{2} \|w\|^2 + \frac{C}{M} L^\nu(w, b, \epsilon) \\ \text{制約} & \epsilon \geq 0 \end{array} \right.$$

と区分的に線形な凸関数の最適化問題として書き換える.

次に, $w = w^0$ と定めた点において, $L^\nu(w, b, \epsilon)$ を下から近似する線形関数を求める. これには, 問題 (4.2) で $w = w^0$ と固定し, ξ, b, ϵ のみを変数とした最適化問題を行えばよい. 定数部分を

$\eta_j^0 = y_j - A_j w^0$ とすれば, 以下の線形計画問題:

$$(4.19) \quad \begin{cases} \text{最小化} & \nu M \epsilon + \sum_{j=1}^M \xi_j \\ \text{制約} & \xi_j - b + \epsilon \geq \eta_j^0, \quad \xi_j + b + \epsilon \geq -\eta_j^0, \quad \xi_j \geq 0, \quad j = 1, 2, \dots, M, \\ & \epsilon \geq 0 \end{cases}$$

が得られる. 双対変数 $\pi \in \mathbb{R}^N$ を定めれば, 双対問題は

$$(4.20) \quad \begin{cases} \text{最大化} & \pi^T \eta^0 \\ \text{制約} & e^T \pi = 0, \quad -e \leq \pi \leq e, \quad e^T |\pi| \leq \nu M \end{cases}$$

となる.

前節と同様に, η^0 の各要素に対して,

$$\eta_1^0 \geq \eta_2^0 \geq \dots \geq \eta_M^0$$

と仮定し, 添え字 t^* を

$$t^* = \lfloor \frac{\nu M}{2} \rfloor$$

と定め, M 次ベクトル h^0 を

$$(4.21) \quad \begin{aligned} h_1^0 &= h_2^0 = \dots = h_{t^*}^0 = 1, \quad h_{t^*+1}^0 = \frac{\nu M}{2} - t^*, \quad h_{t^*+2}^0 = h_{t^*+3}^0 = \dots = h_{M-t^*-1}^0 = 0, \\ h_M^0 &= h_{M-1}^0 = \dots = h_{M-t^*+1}^0 = -1, \quad h_{M-t^*}^0 = -\frac{\nu M}{2} + t^* \end{aligned}$$

と定義する.

定理 4.5. h^0 を (4.21) で定められた M 次の 0-1 ベクトルとすれば, $\pi = h^0$ は問題 (4.20) の最適解である.

証明 h^0 が双対問題 (4.20) の実行可能解であることは, 容易に確かめられる.

一方, b^0, ϵ^0 を次の 2 式

$$\begin{aligned} \xi_{t^*+1}^0 &= \eta_{t^*+1}^0 + b^0 - \epsilon^0 \\ \xi_{M-t^*}^0 &= -\eta_{M-t^*}^0 - b^0 - \epsilon^0 \end{aligned}$$

を満たすように定める. すなわち,

$$b^0 = \frac{-\eta_{t^*+1}^0 - \eta_{M-t^*}^0}{2}, \quad \epsilon^0 = \frac{\eta_{t^*+1}^0 - \eta_{M-t^*}^0}{2}$$

と定め, M 次ベクトル ξ^0 を

$$\begin{aligned} \xi_j^0 &= \eta_j^0 + b^0 - \epsilon^0, \quad j = 1, 2, \dots, t^* \\ \xi_j^0 &= 0, \quad j = t^* + 1, t^* + 1, \dots, M - t^*, \\ \xi_j^0 &= -\eta_j^0 - b^0 - \epsilon^0, \quad j = M - t^* + 1, \dots, M - 1, M \end{aligned}$$

とする. このように定めれば, (ξ^0, b^0, ϵ^0) は問題 (4.19) の実行可能解である.

主問題と双対問題の目的関数が一致していることより, 最適性が示される. □

よって以下の不等式が成立する.

補題 4.6. 任意のベクトル $w^0 \in \mathbb{R}^n$ に対応して定まる問題 (4.20) の最適解を h^0 とする. このとき, 不等式

$$L^\nu(w, b, \epsilon) \geq h^{0T} y - h^{0T} A w$$

が成立する. また, b^0, ϵ^0 を (4.19) の最適解とすれば, $(w, b, \epsilon) = (w^0, b^0, \epsilon^0)$ において等号が成立する.

そこで, s 個のベクトル h^k ($k = 1, 2, \dots, s$) が得られたと仮定すれば, 次の最適化問題:

$$(4.22) \quad \begin{cases} \text{最小化} & \frac{1}{2} \|w\|^2 + \frac{C}{M} r \\ \text{制約} & r \geq h^{kT} y - h^{kT} A w, \quad k = 1, 2, \dots, s, \end{cases}$$

を解くことで問題 (4.2) の下界が得られる. さらに,

$$H^s = [h^1 \ h^2 \ \dots \ h^s]$$

と定義すれば, (4.22) の双対問題は

$$(4.23) \quad \begin{cases} \text{最小化} & \frac{1}{2} \beta^T H^{sT} A A^T H^s \beta - y^T H^s \beta \\ \text{制約} & e^T \beta = \frac{C}{M}, \quad \beta \geq 0 \end{cases}$$

である. また, β^* を双対問題の最適解とすれば, 主問題の最適解 w^* は

$$w^* = A^T H^s \beta^*$$

と求められる. なお, ϵ -SVR のときと同様に, カーネルを用いた非線形な回帰を行う場合には, $A A^T$ の部分をカーネル行列 K に置き換え,

$$(4.24) \quad \begin{cases} \text{最小化} & \frac{1}{2} \beta^T H^{sT} K H^s \beta - y^T H^s \beta \\ \text{制約} & e^T \beta = \frac{C}{M}, \quad \beta \geq 0 \end{cases}$$

を最適化すればよい. 以上のアルゴリズムをまとめると, 図 5 となる.

参考文献

- [1] C.-C. CHANG AND C.-J. LIN, *LIBSVM: a library for support vector machines*, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [2] R. COLLOBERT AND S. BENGIO, *SVM-Torch: Support vector machines for large-scale regression problems*, Journal of Machine Learning Research, 1 (2001), pp. 143–160. Software available at <ftp://idiap.ch/pub/learning/SVM-Torch.tgz>.
- [3] C. CORTES AND V. VAPNIK, *Support-vector networks*, Machine learning, 20 (1995), pp. 273–297.
- [4] T. JOACHIMS, *Text categorization with support vector machines: Learning with many relevant features*, Lecture notes in computer science, 1398 (1998), pp. 137–142.

- [5] ———, *Making large-scale support vector machine learning practical*, in Advances in Kernel Methods, B. Schölkopf, C. Burges, and A. Smola, eds., The MIT Press, 1999, pp. 169–184. Software available at <http://svmlight.joachims.org/>.
- [6] O. L. MANGASARIAN, *Nonlinear Programming*, vol. 10 of Classics in Applied Mathematics, SIAM, Philadelphia, PA, 1994.
- [7] B. SCHÖLKOPF, C. BURGESS, AND V. VAPNIK, *Extracting support data for a given task*, in Proceedings, First International Conference on Knowledge Discovery & Data Mining, U. M. Fayyad and R. Uthurusamy, eds., 1995, pp. 252–257.
- [8] B. SCHÖLKOPF, A. SMOLA, R. WILLIAMSON, AND P. BARTLETT, *New support vector algorithms*, Neural Computation, 12 (2000), pp. 1207–1245.
- [9] R. J. VANDERBEI, *LOQO: An interior point code for quadratic programming*, Optimization Methods and Software, 11 (1999), pp. 451–484.
- [10] V. N. VAPNIK, *The nature of statistical learning theory*, Statistics for Engineering and Information Science, Springer-Verlag, New York, 2000.